DOCUMENT RESUME

ED 466 774		TM 034 285

AUTHOR	Lowe, Terry J.
TITLE	Use and Interpretation of Effect Sizes and Structure Coefficients in Mathematics and Science Education Journals.
PUB DATE	2002-02-15
NOTE	16p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, February 14-16, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Effect Size; *Mathematics Education; Research Methodology; *Research Reports; *Scholarly Journals; *Science Education; *Sciences
IDENTIFIERS	*Structure Coefficients

ABSTRACT

Reporting and interpretation of effect sizes and structure coefficients in multiple regression results are important for good practice. The purpose of this study was to investigate the use and interpretation of effect sizes (ES) and structure coefficients in multiple regression analyses in two mathematics and science education journals. Published studies in five complete volumes of the "Journal for Research in Mathematics Education" and "School Science and Mathematics" were inspected. Nineteen multiple regression research articles were located and coded. Results suggest that effects are generally reported by researchers using multiple regression, but structure coefficients are seldom examined to help evaluate variable importance. (Contains 41 references.) (Author/SLD)

Running head: EFFECT SIZES AND STRUCTURE COEFFICIENTS

Use and Interpretation of Effect Sizes and Structure Coefficients in Mathematics and Science Education Journals

Terry J. Lowe

University of North Texas

Paper presented at the annual meeting of the Southwest Educational Research Association, February 15, 2002, Austin, TX.

Abstract

Reporting and interpretation of effect sizes and structure coefficients in multiple regression results is important for good practice. The purpose of this paper is to investigate the use and interpretation of effect size (ES) and structure coefficients in multiple regression analyses in two mathematics and science education journals. Published studies in the Journal for Research in Mathematics Education and School Science and Mathematics were inspected. Results suggest that effects are generally reported by researchers using multiple regression, but structure coefficients are seldom examined to help evaluate variable importance.

Use and Interpretation of Effect Sizes and Structure Coefficients in Mathematics and Science Education Journals

Multiple regression analysis is a common tool used in social science research. Multiple regression can be useful in both predictive and explanatory research (Pedhazur, 1997). More than three decades ago, Jacob Cohen (1968) demonstrated that multiple regression was the univariate general linear model (GLM) and subsumed other univariate methods as special cases. Knapp (1978) later demonstrated canonical correlation as the multivariate GLM (see also Bagozzi, Fornell & Larcker, 1981; Henson, 2000; Thompson 1991).

Because all GLM analyses are correlational in nature, they all yield $\underline{r}^2$-type effect sizes that should be both reported and interpreted (cf. Cohen, 1994; Henson & Smith, 2000; Thompson, 1996). They also invoke weights (e.g., beta weights) that are applied to observed variables to create synthetic variables (e.g. Yhat predicted scores) that become the focus of the analysis. Therefore, the relationship between the observed variables and the synthetic variables becomes important in determining the value of the observed variable. The correlation between an observed and synthetic variable is called a structure coefficient, which are also present in all GLM analyses (Courville & Thompson, 2001; Thompson & Borrello, 1985). In regression, for example, the correlation (structure coefficient) between a predictor and the Yhat predicted scores informs the researcher as to the potential contributory value of the predictor to the effect observed.

Further, within any GLM analysis such as multiple regression, researchers are typically faced with a two-stage process of result interpretation. Stage one concerns whether the researcher has a result that is noteworthy, which historically has been evaluated with statistical significance testing. More recently, the role and import of effect sizes have seen increased use, including a mandate to "always" report effects because such reporting is "essential" to good research practice (Wilkinson & APA Task Force on Statistical Inference, 1999, p. 599).

Once a noteworthy effect is observed, then (and only then), the researcher likely is concerned with determining what variables contributed to the effect. This is typically done by looking at standardized weights, but should also include examination of structure coefficients. For example, in multiple regression, a noteworthy $\underline{R}^2$ effect is generally followed by study of the predictor beta weights (or unstandardized b weights) but should also invoke examination of the predictor structure coefficients. Courville and Thompson (2001) provide a thorough discussion on the definition and role of structure coefficients in multiple regression. Henson (2002) discusses structure coefficients in multivariate analyses.

Purpose

The purpose of the present paper is to investigate if suggested reporting standards of statistical information are being followed in two mathematics and science education journals. Focus was on the use and interpretation of effect size and structure coefficients in multiple regression applications. The literature suggests effect size reporting is generally sparse (DeVaney, 2001; Henson & Smith, 2000; Thompson, 2001), and the interpretation of effect sizes occurs even less often. However, Kirk (1996) observed higher rates of reporting for journals that frequently published articles employing multiple regression. Kirk postulated that effect reporting was greater for these journals because statistical software packages routinely provide $\underline{R}^2$ effects in their outputs. This begs the question of whether researchers are thoughtfully including effects or if reporting is a function of what the computer output provides. Although an effect is reported, whether the effect is interpreted is another matter altogether.

Further, the reporting and interpretation of structure coefficients tends to be the exception rather than the rule. For the present study, I expected to observe moderate frequency of effect size reporting and very minimal use of structure coefficients.

Statistical Significance Testing and Effect Size

Statistical significant testing became common practice in the early to mid-1900's with its introduction into textbooks (DeVaney, 2001). Fan (2001) stated "..., statistical significance testing answers the question: What is the probability of obtaining an observed sample statistic ... when the population has a known parameter value ...?" (p. 275). Statistical significance testing is studied in most statistics courses, which often leaves graduate students with the idea that statistical significance testing is the most important part of the statistical interpretation. Further, the common p-value is routinely the only statistic interpreted in educational research (Plucker, 1997).

Statistical significance testing has supporters and critics. The supporters take the position that when used correctly, statistical significance testing describes valuable information (DeVaney, 2001; Henson & Smith, 2000). Critics express that statistical significance testing is often misinterpreted as practical significance (DeVaney, 2001; Henson & Smith, 2000). The dependence of statistical significance testing on sample size often confuses correct interpretations of statistical significance (DeVaney, 2001; Fan, 2001; Plucker, 1997). With the apparent limitations of statistical significance tests, effect sizes have been advocated for result interpretation (Fan, 2001).

Effect size in multiple regression typically refers to a variance-accounted-for statistic known as R^2. Effect size does not replace statistical significance testing, and effect size reporting should complement statistical significance testing. After reporting the results of significance testing, effect sizes should be reported and interpreted. Plucker (1997) reported that researchers in the social sciences are encouraged to provide effect size or the information for the calculation of the effect size in the results of studies. The fifth edition of the APA Publication Manual expects the reporting of effect size (APA, 2001). However, as Henson and Smith (2000) noted, "A more appropriate measure of effect size use would include an assessment of whether researchers both report and interpret their obtained effect sizes" (p. 290, emphasis added).

Henson and Smith (2000) suggested that the trend toward reporting of effect size is progressing slowly. Even with the encouragement to report effect size by fourth edition of the APA Manuel (APA, 1994), the reporting of effect size is often neglected (DeVaney, 2001). Eleven empirical studies have been completed that show this encouragement by the APA has been ineffective (Thompson, 2001).

Therefore, one purpose of this research seeks to determine the extent to which authors of mathematics and science education studies report and interpret effect size.

Beta Weights and Structure Coefficients

After finding noteworthy effects, the researcher is charged with determining the variables that contributed to those effects. In regression and other GLM analyses, this determination should invoke study of both standardized weights and structure coefficients.

Beta weights are standardized coefficients (i.e., regression coefficients generated from standardized data) that suggest the independent variable's predictive contribution to the dependent variable. Beta weights are easy to determine as they are generated in the output of multiple regression computer programs. However, beta weights are affected by collinearity, when predictor variables are correlated with each other. Hence, beta weights may yield distorted interpretations of variable import (Thompson & Borrello, 1985).

Structure coefficients in multiple regression are the correlations between the predictor variable and the estimated (predicted) score in the regression analysis (i.e., Yhat). Structure coefficients (in squared format) express how much variance in the predicted variable (Yhat) that the predictor variable possibly could explain. Because they are simple bivariate correlations, structure coefficients are not affected by collinearity as betas are (Thompson & Borrello, 1985).

Because beta weights may not express accurate variable interpretation, structure coefficients are being supported as an important step in statistical interpretation (Courville & Thompson, 2000; Thompson , 2001). Cooley and Lohnes (Thompson, 1990) argue for the reporting of structure coefficients by the general researcher. Of course, structure coefficients yield the same interpretation as zero-order correlations between the predictor and the dependent variable (Knapp, 1978). However, structures are interpretation in terms of the synthetic Yhat variable.

Method

Multiple regression studies obtained from the Journal for Research in Mathematics Education and the School Science and Mathematics journal were identified. Every journal article of each issue was inspected for articles using multiple regression. I began with the latest issue and continued back until five complete volumes of each journal were exhausted. For the Journal of Research in Mathematics Education, I began with Volume 32, Number 4 and stopped after the inspection of Volume 27. For School Science and Mathematics, I began with Volume 101, Number 7 and stopped after completing volume 96. After a multiple regression analysis was located, the article was read for the reporting and interpretation of effect sizes, Pearson correlation coefficients (r), and structure coefficients. This information was coded in a table. Also, the type of effect size was noted.

The first multiple regression from each article was selected for inspection. A study received credit for reporting the various information if the information was described in tables or in words. A study received credit for interpretation of effect sizes, Pearson correlation coefficient, and structure coefficients if an explicit description of or reference to the statistic was included.

Results

Nineteen multiple regression research articles were located and coded. Simple frequencies for observations are presented in Table 1.

Table 1.

Summary of Reporting and Interpretation in Multiple Regression Articles.

	Report			Interpret		
	Effect Size	Pearson r	Structure Coefficients	Effect Size	Pearson r	Structure Coefficients
Yes	16 (84%)	8 (42%)	1 (5%)	16 (84%)	9 (47%)	1 (5%)
No	3 (16%)	11 (58%)	18 (95%)	3 (16%)	10 (53%)	18 (95%)
Total	19	19	19	19	19	19

Approximately 84% of the articles reviewed reported effect size. The interpretation of the effect size was the same at approximately 84% of the studies. Every article that reported effect size, interpreted the effect size. Three studies did not report or interpret effect size. When the effect size was reported, $\underline{R}^2$ was the statistic of choice in all cases.

Pearson correlation coefficients ($\underline{r}$) were reported in approximately 42% of the multiple regression articles. All articles that reported the Pearson correlation coefficient, also interpreted the coefficient. One study interpreted Pearson correlation coefficients but did not report the statistic (Schwartz, 2000).

The use of structure coefficients was as expected and almost non-existent. Only one multiple regression article reported and interpreted structure coefficients (Terry & Baird, 1997).

Discussion

These results of the investigation of multiple regression analyses in mathematics and science education journals disagree with other studies that suggest the reporting of effect size to be slowly occurring (Henson & Smith, 2000) and often neglected (DeVaney, 2001). When using multiple regression analyses, mathematics and science researchers appear to have been responsive to the recommendations of the APA Publication Manual (APA, 2001). However, as Kirk (1996) suggested effect reporting is likely higher for regression research due to the commonly known $\underline{R}^2$ effect size routinely provided by statistical software outputs.

In this review of the literature, the only effect size reported was $\underline{R}^2$. This is consistent with a report by Fan (2001) that $\underline{R}^2$ is the most popular effect size reported in several journals of psychology and Kirk's (1996) review. Statistical software packages routinely present this information in their regression output, making $\underline{R}^2$ readily available. Therefore, when regression is used, effect reporting tends to be more frequent. However, for analyses that statistical packages tend not to output effects, reporting is more sparse. This reality begs the question of whether effect reporting is a function of thoughtful researcher judgment or statistical output.

Adjusted $\underline{R}^2$ is also printed in the regression output of most statistical packages. However, adjusted $\underline{R}^2$ was not reported in the current study. Adjusted $\underline{R}^2$ is a "corrected" effect and will be less than or equal to the $\underline{R}^2$. Corrected effects correct for sampling error and are better estimates of the true population effect.

Unfortunately, mathematics and science education researchers seldom examine structure coefficient to determine predictor importance. Zero-order correlations are also seldom interpreted beyond beta weights or unstandardized coefficients. Because statistical packages do not routinely report these coefficients in their outputs, structure coefficients are not as easily found (although they may be found by simply dividing the bivariate correlation between a predictor and the dependent variable by the multiple $\underline{R}$). Because of this lack of reporting and interpretation, it is certainly possible that the researchers made erroneous conclusions about what predictors could have predicted the dependent variable. Courville and Thompson (2001) provide several examples of such misstatements.

In sum, good research multiple regression practice should consider effect sizes when deciding if a result is noteworthy and both standardized weights and structure coefficients when deciding what predictors were valuable in generating the observed effect. The present review of two mathematics and science education journals suggest reasonable reporting of effects, with an equal amount of interpretation of those effects, and almost non-existent examination of structure coefficients.

References

American Psychological Association. (1994). Publication manual of the American Psychological Association. Washington DC: Author.

American Psychological Association. (2001). Publication manual of the American Psychological Association. Washington DC: Author.

Bagozzi, R. P., Fornell, C., & Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. Multivariate Behavioral Research, 16, 437-454.

*Carr, M., Jessup, D. L., & Fuller, D. (1999). Gender differences in first-grade mathematics strategy use: Parent and teacher contributions. Journal for Research in Mathematics Education, 30(1), 20-46.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, J. (1994). The earth is round (p<.05). American Psychologist, 49, 997-1003.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. Educational and Psychological Measurement, 61, 229-248.

*Czerniak, C. M., & Lumpe, A. J. (1996). Predictors of science fair participation using the theory of planned behavior. School Science & Mathematics, 96(7), 355-361.

DeVaney, T. A. (2001). Statistical significance, effect size, and replication: What do the journals say? Journal of Experimental Education, 69(3), 310-319.

*Duncan, H., & Dick, T. (2000). Collaborative workshops and student academic performance in introductory college mathematics courses: A study of a treisman model math excel program. School Science & Mathematics, 100(7), 365-373.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. The Journal of Educational Research, 94(5), 275-282.

*Gearhart, M., Saxe, G. B., Seltzer, M., Schlackman, J., Ching, C. C., Nasir, N., Fall, R., Bennett, T., Rhine, S., & Sloan, T. F. (1999). Opportunities to learn fractions in elementary mathematics classrooms. Journal for Research in Mathematics Education, 30(3), 286-315.

*Gfeller, M. K., Niess, M. L., & Lederman, N. G. (1999). Preservice teacher's use of multiple representations in solving arithmetic mean problems. School Science & Mathematics, 99(5), 250-257.

Henson, R.K. (2000). Demystifying parametric analyses: Illustrating canonical correlation as the multivariate general linear model. Multiple Linear Regression Viewpoints, 26(1), 11-19.

Henson. R. K. (2002, April). The logic and interpretation of structure coefficients in multivariate general linear model analyses. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Reproduction Service No. forthcoming)

Henson, R.K., & Smith, A.D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force report and current trends. Journal of Research and Development in Education, 33(4). 285-296.

*Joyce, B. A., & Farenga, S. (1999). Informal science experience, attitudes, future interest in science, and gender of high-ability students: An exploratory study. School Science & Mathematics, 99(8), 431-437.

Kirk, R.E. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement, 56, 746-759.

Knapp, T.R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Knapp, T.R., & Sawilowsky, S. S. (2001). Strong arguments: Rejoiner to Thompson. Journal of Experimental Education, 70(1), 94-95.

*Lumpe, A., & Haney, J. J. (1998). Science teacher beliefs and intentions regarding the use of cooperative learning. School Science & Mathematics, 98(3), 123-132.

*Ma, X. (1999). A meta-analysis of the relationship between anxiety towards mathematics and achievement. Journal for Research in Mathematics Education, 30(5), 520-540.

*Morrell, P. D., & Lederman, N. G. (1998). Students' attitudes toward school and classroom science: Are they independent phenomena? School Science and Mathematics, 98(2), 76-86.

*Paek, P., & Holland, P. (1999). Development and analysis of a mathematics aptitude test for gifted elementary school students. School Science & Mathematics, 99(6), 338-347.

Pedhazur, E. J. (1997). Multiple regression in behavioral research: Explanation and prediction. Fort Worth, TX: Harcourt College Publishers.

*Pesek, D.D. (2000). Interference of instrumental instruction in subsequent relational learning. Journal for Research in Mathematics Education, 31(5), 524.

Plucker, J.A. (1997). Debunking the myth of the 'highly significant' result: Effect sizes in gifted education research. Roeper Review, 20(2), 122-126.

*Ramos, I., & Lambating, J. (1996). Gender differences in risk-taking behavior and their relationship to SAT-mathematics performance. School Science & Mathematics, 96(4), 202-207.

*Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. Journal for Research in Mathematics Education, 32(4), 368-398.

*Schwartz, R. (2000). Achieving the reforms vision: The effectiveness of a specialists-led elementary science program. School Science & Mathematics, 100(4), 181-193.

*Shepardson D. P. (1997). The nature of student thinking in life science laboratories. School Science & Mathematics, 97(1), 37-43.

*Tarr, J. E., Uekawa, K., Mittag, K. C., & Lennex, L. (2000). A comparison of calculator use in eighth-grade mathematics classrooms in the United States, Japan, and Portugal: results from the Third International Mathematics and Science Study. School Science & Mathematics, 100(3), 139-150.

*Terry J. M., & Baird, W. E. (1997). What factors affect attitudes toward women in science held by high school biology students? School Science & Mathematics, 97(2). 78-86.

Thompson, B. (1990). Don't forget the structure coefficients. Measurement and Evaluation in Counseling and Development, 22(4), 178-180.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24, 80-95.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. Journal of Experiential Education, 70(1), 80-93.

Thompson, B., & Borrello, G. A. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

*Valanides, N. C. (1996). Formal reasoning and science teaching. School Science & Mathematics, 96(2), 99-107.

*Wang, J., Wildman, L., & Calhoun, G. (1996). The relationship between parental influence and student achievement in seventh grade mathematics. School Science & Mathematics, 96(8), 395-399.

Wilkinson, L. & APA Task Force on Statistical Inference. (1999). Statistical Methods in psychology journals: Guidelines and explanation. American Psychologist, 54, 594-604.

TM034285

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Use and Interpretation of Effect Sizes and Structure Coefficients in Mathematics and Science Journals	
Author(s): Terry J. Lowe	
Corporate Source: University of North Texas	Publication Date: February 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Terry J. Lowe	Printed Name/Position/Title: Terry J. Lowe / Research Associate	
Organization/Address: 922 Highland Dr Cleburne TX 76033	Telephone: (817) 645-4114	FAX:
	E-Mail Address: terrlowe@swau.edu	Date: 2/6/02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com